Roda et al_tips

install.packages("zoo")

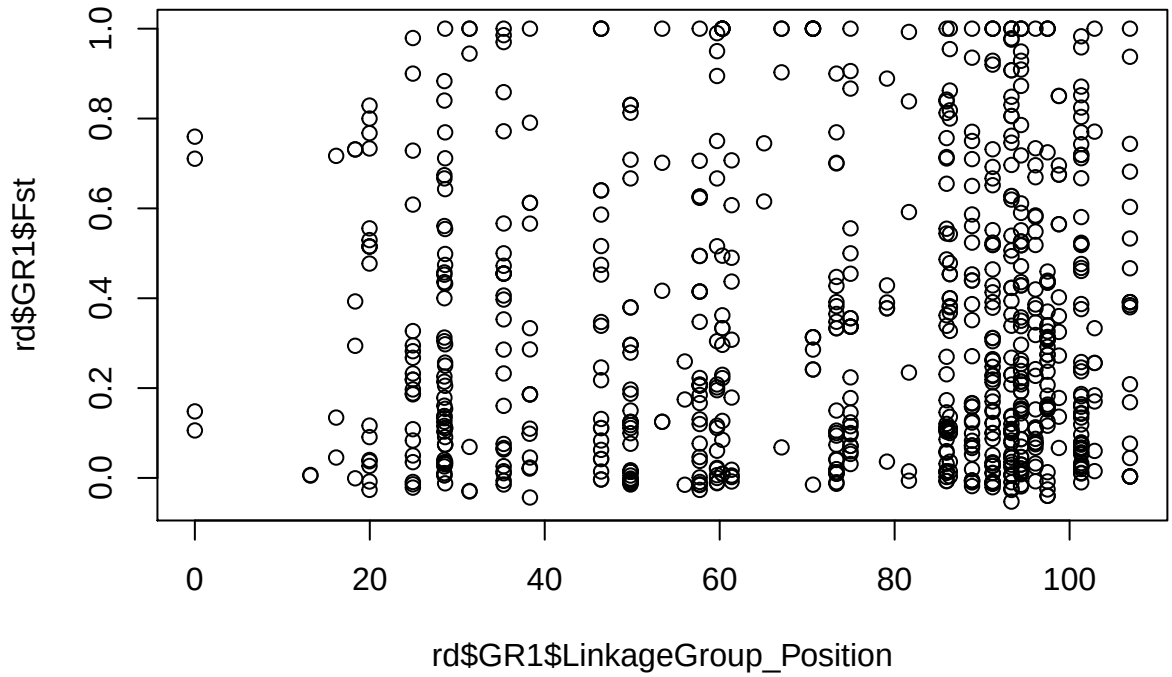
Tips for the Roda et al data:

```
#Getting started:
library(ggplot2)
roda<-read.csv("data/Fst_BSA_wLinkageGrp.csv", header=T)
head(roda) #will show you the first 5 rows of each column in the dataset
```

```
##   LocusID Length Coverage LinkageGroup Locus_Position
## 1 1000527   4893  9.05886         GR11             689
## 2 1000527   4893  9.05886         GR11             641
## 3 1000527   4893  9.05886         GR11             606
## 4 1000527   4893  9.05886         GR11             751
## 5 1000527   4893  9.05886         GR11             751
## 6 1000527   4893  9.05886         GR11             751
##                       Comparison maa_pop1 maa_pop2 Count_maa_pop1
## 1 F8-C-S-Dune_vs_F8-C-S-Headland        T        C            134
## 2 F8-C-S-Dune_vs_F8-C-S-Headland        G        A            134
## 3 F8-C-S-Dune_vs_F8-C-S-Headland        A        C            134
## 4 F8-A-S-Dune_vs_F8-A-S-Headland        G        A             45
## 5 F8-C-S-Dune_vs_F8-C-S-Headland        A        A            106
## 6 F8-B-S-Dune_vs_F8-B-S-Headland        G        G             26
##   Count_Total_pop1 Count_maa_pop2 Count_Total_pop2        Fst  P.value_FET
## 1              135             42               42 0.99253731 4.640000e-40
## 2              135             42               42 0.99253731 4.640000e-40
## 3              135             33               42 0.76982017 1.960000e-26
## 4               76             34               34 0.58666667 1.190000e-10
## 5              106             92              106 0.12380952 3.860000e-05
## 6               47             81              106 0.07928435 5.382402e-03
##   Outlier_FET_Fst PositionMethod LinkageGroup_Position
## 1               1          noisy                44.832
## 2               1          noisy                44.832
## 3               0          noisy                44.832
## 4               0          noisy                44.832
## 5               0          noisy                44.832
## 6               0          noisy                44.832
```

```
roda<-roda[order(roda$LinkageGroup, roda$LinkageGroup_Position),]
#We want to look at each linkage group at a time:
rd<-split(roda, roda$LinkageGroup)
```

```
#mapping all values:
plot(rd$GR1$LinkageGroup_Position, rd$GR1$Fst)
```




```
#it's really hard to see what's going on here- just looks like a big block of data. Let's try creating

#we have to call our windowing function "SWsnp"
library(zoo)
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
SWsnp = function(x,y,z){
  out<-rollapply(as.numeric(x),width=y,mean,by= z, na.rm=TRUE)
}
```

```
#the window size can be somewhat arbitrary- it will depend on how many SNPs you have. Making bigger win
LG1_pos<-SWsnp(rd$GR1$LinkageGroup_Position, 20,2)
LG1_Fst<-SWsnp(rd$GR1$Fst, 20,2)
LG1_LG<-rep("LG1",length(LG1_pos))

LG2_pos<-SWsnp(rd$GR2$LinkageGroup_Position, 20,2)
LG2_Fst<-SWsnp(rd$GR2$Fst, 20,2)
LG2_LG<-rep("LG2",length(LG2_pos))

LG3_pos<-SWsnp(rd$GR3$LinkageGroup_Position, 20,2)
LG3_Fst<-SWsnp(rd$GR3$Fst, 20,2)
LG3_LG<-rep("LG3",length(LG3_pos))

LG4_pos<-SWsnp(rd$GR4$LinkageGroup_Position, 20,2)
LG4_Fst<-SWsnp(rd$GR4$Fst, 20,2)
LG4_LG<-rep("LG4",length(LG4_pos))

LG5_pos<-SWsnp(rd$GR5$LinkageGroup_Position, 20,2)
```

```
LG5_Fst<-SWsnp(rd$GR5$Fst, 20,2)
LG5_LG<-rep("LG5",length(LG5_pos))

LG6_pos<-SWsnp(rd$GR6$LinkageGroup_Position, 20,2)
LG6_Fst<-SWsnp(rd$GR6$Fst, 20,2)
LG6_LG<-rep("LG6",length(LG6_pos))

LG7_pos<-SWsnp(rd$GR7$LinkageGroup_Position, 20,2)
LG7_Fst<-SWsnp(rd$GR7$Fst, 20,2)
LG7_LG<-rep("LG7",length(LG7_pos))

LG8_pos<-SWsnp(rd$GR8$LinkageGroup_Position, 20,2)
LG8_Fst<-SWsnp(rd$GR8$Fst, 20,2)
LG8_LG<-rep("LG8",length(LG8_pos))

LG9_pos<-SWsnp(rd$GR9$LinkageGroup_Position, 20,2)
LG9_Fst<-SWsnp(rd$GR9$Fst, 20,2)
LG9_LG<-rep("LG9",length(LG9_pos))

LG10_pos<-SWsnp(rd$GR10$LinkageGroup_Position, 20,2)
LG10_Fst<-SWsnp(rd$GR10$Fst, 20,2)
LG10_LG<-rep("LG10",length(LG10_pos))

LG11_pos<-SWsnp(rd$GR11$LinkageGroup_Position, 20,2)
LG11_Fst<-SWsnp(rd$GR11$Fst, 20,2)
LG11_LG<-rep("LG11",length(LG11_pos))

LG12_pos<-SWsnp(rd$GR12$LinkageGroup_Position, 20,2)
LG12_Fst<-SWsnp(rd$GR12$Fst, 20,2)
LG12_LG<-rep("LG12",length(LG12_pos))

LG13_pos<-SWsnp(rd$GR13$LinkageGroup_Position, 20,2)
LG13_Fst<-SWsnp(rd$GR13$Fst, 20,2)
LG13_LG<-rep("LG13",length(LG13_pos))

LG14_pos<-SWsnp(rd$GR14$LinkageGroup_Position, 20,2)
LG14_Fst<-SWsnp(rd$GR14$Fst, 20,2)
LG14_LG<-rep("LG14",length(LG14_pos))

LG15_pos<-SWsnp(rd$GR15$LinkageGroup_Position, 20,2)
LG15_Fst<-SWsnp(rd$GR15$Fst, 20,2)
LG15_LG<-rep("LG15",length(LG15_pos))

LG16_pos<-SWsnp(rd$GR16$LinkageGroup_Position, 20,2)
LG16_Fst<-SWsnp(rd$GR16$Fst, 20,2)
LG16_LG<-rep("LG16",length(LG16_pos))

LG17_pos<-SWsnp(rd$GR17$LinkageGroup_Position, 20,2)
LG17_Fst<-SWsnp(rd$GR17$Fst, 20,2)
LG17_LG<-rep("LG1",length(LG17_pos))

LG18_pos<-SWsnp(rd$GR18$LinkageGroup_Position, 20,2)
LG18_Fst<-SWsnp(rd$GR18$Fst, 20,2)
```

```
LG18_LG<-rep("LG1",length(LG18_pos))

LG19_pos<-SWsnp(rd$GR19$LinkageGroup_Position, 20,2)
LG19_Fst<-SWsnp(rd$GR19$Fst, 20,2)
LG19_LG<-rep("LG1",length(LG19_pos))

LG20_pos<-SWsnp(rd$GR20$LinkageGroup_Position, 20,2)
LG20_Fst<-SWsnp(rd$GR20$Fst, 20,2)
LG20_LG<-rep("LG1",length(LG20_pos))
```

```
#putting all the windowed values back in 1 dataframe:
LG_names<-rbind(LG1_LG, LG2_LG, LG3_LG, LG4_LG, LG5_LG, LG6_LG, LG7_LG, LG8_LG, LG9_LG, LG10_LG, LG11_LG
```

```
## Warning in rbind(LG1_LG, LG2_LG, LG3_LG, LG4_LG, LG5_LG, LG6_LG, LG7_LG, :
## number of columns of result is not a multiple of vector length (arg 1)
```

```
LG_pos<-rbind(LG1_pos, LG2_pos, LG3_pos, LG4_pos, LG5_pos, LG6_pos, LG7_pos, LG8_pos, LG9_pos, LG10_pos
```

```
## Warning in rbind(LG1_pos, LG2_pos, LG3_pos, LG4_pos, LG5_pos, LG6_pos,
## LG7_pos, : number of columns of result is not a multiple of vector length
## (arg 1)
```

```
LG_fst<-rbind(LG1_Fst, LG2_Fst, LG3_Fst, LG4_Fst, LG5_Fst, LG6_Fst, LG7_Fst, LG8_Fst, LG9_Fst, LG10_Fst
```

```
## Warning in rbind(LG1_Fst, LG2_Fst, LG3_Fst, LG4_Fst, LG5_Fst, LG6_Fst,
## LG7_Fst, : number of columns of result is not a multiple of vector length
## (arg 1)
```

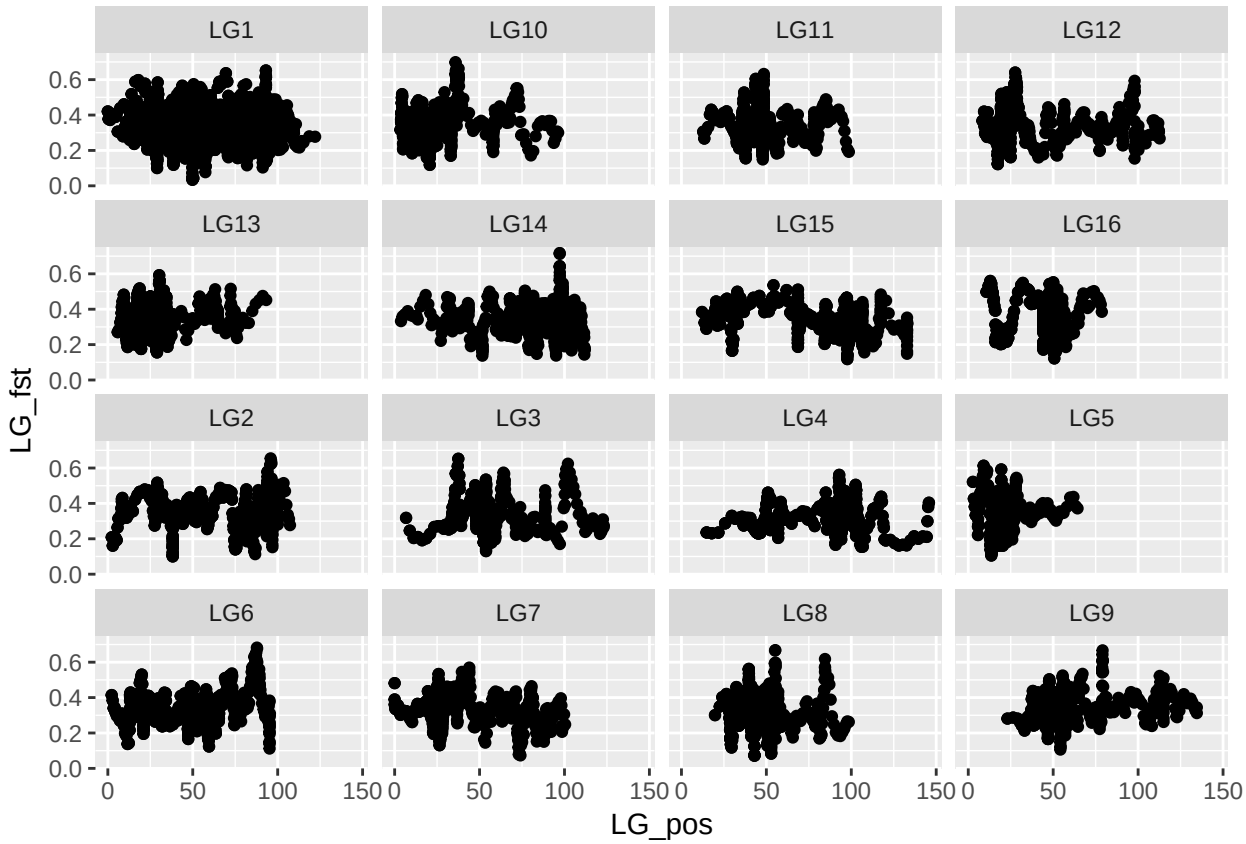
```
LG_names<-as.vector(LG_names)
LG_pos<-as.vector(LG_pos)
LG_fst<-as.vector(LG_fst)

data<-cbind(LG_names, LG_pos, LG_fst)
data<-as.data.frame(data)
```

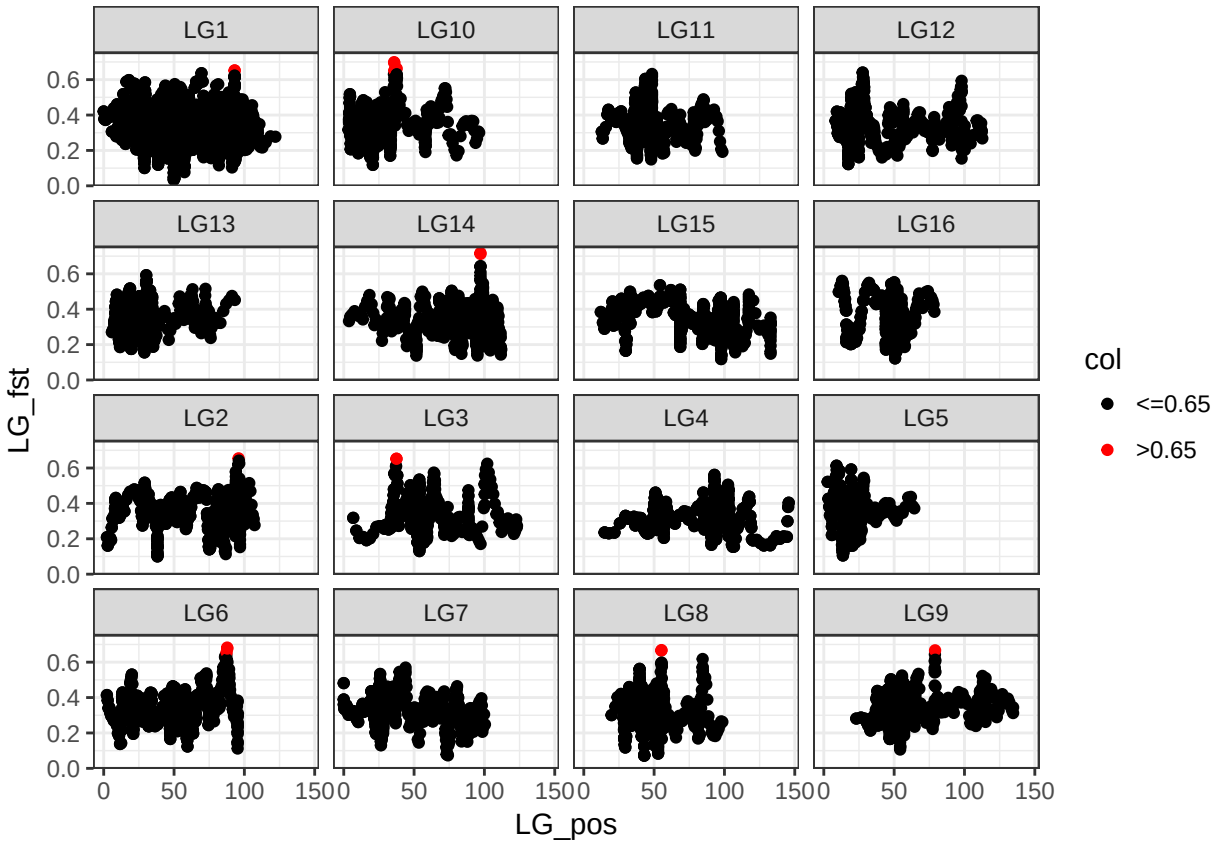
```
# we've coerced the data a few times, so we need to reformat our numeric data in order for it to plot i


data[[2]] <- as.numeric(as.character(data[[2]]))
data[[3]] <- as.numeric(as.character(data[[3]]))

#Let's plot it now:
ggplot(data, aes(x=LG_pos, y=LG_fst))+geom_point()+facet_wrap(~LG_names)
```




```
#lets plot some high Fst values in another color. We'll use 0.65 as a cutoff
data$col <- cut(data$LG_fst,
                breaks = c(-Inf, 0.65, Inf),
                labels = c("<=0.65", ">0.65"))
ggplot(data, aes(x=LG_pos, y=LG_fst, color=col))+ geom_point()+facet_wrap(~LG_names)+scale_color_manual
```




```
# try plotting another variable- perhaps the outlier status!
# try plotting the different comparisons
# Want to plot something totally different? feel free to change everything, as long as you think it tel
```